Evolutionary rates across gradients

Jason T. Weir* and Adam Lawson

Department of Biological Sciences and Department of Ecology and Evolutionary Biology, University of Toronto Scarborough, Toronto, ON, M1C 1A4, Canada

Summary

1. A key question in the fields of macroecology and evolution is how rates of evolution vary across gradients, be they ecological (e.g. temperature, rainfall, net primary productivity), geographic (e.g. latitude, elevation), morphological (e.g. body mass), etc. Evolutionary rates across gradients (EVRAG 2.0) is a new software package provided as open source in the \textit{R} language (and available from CRAN) that tests whether rates of trait evolution vary continuously across such continuous gradients.

2. The approach uses quantitative trait data for a series of sister-pair contrasts (i.e. sister species or other types of sister taxa) and applies Brownian Motion and Ornstein Uhlenbeck models in which parameter (evolutionary rate and constraint) values vary as a function of discrete variables (e.g. male vs. female) and/or continuous variables (e.g. latitude, temperature, body size).

3. We used simulation to test performance of the models in EvoRAG. Our gradient models accurately estimate parameter values, have very low levels of bias, and low rates of model misspecification.

4. The modelling framework developed here provides great flexibility in designing models that test how rates of evolution vary across gradients. We provide an example where both discrete (songbird vs. subsong) and continuous (latitudinal) effects on evolutionary rate in avian song were simultaneously estimated.

Key-words: bird song, Brownian Motion, continuous character, evolutionary rates, gradients, Ornstein Uhlenbeck, song learning

Introduction

Investigation of the rates and patterns of trait evolution along dated molecular phylogenies is now common thanks to the development of a number of packages in the \textit{R} and other languages (e.g. auteur, Eastman et al. 2011; ape, Paradis, Claude & Strimmer 2004; BANNM, Rabosky 2014; geiger, Harmon et al. 2008; motmot, Thomas & Freckleton 2012; OUCH, King & Butler 2009; OUwie, Beaulieu & O’Meara 2012; phyltools, Revell 2012; RBrownie, Stack, Harmon & O’Meara 2011; and others). Many studies investigating trait evolution aim to uncover how evolutionary rates have changed through the radiation of a focal clade by testing for slowdowns in rates that might be indicative of adaptive radiation (e.g. Harmon et al. 2006; Rabosky 2010; Thomas & Freckleton 2012; Revell et al. 2012; Rabosky 2014). One method (Brownie, O’Meara et al. 2006) allows separate evolutionary rates to be applied to species that occur in different discrete categorical states independent of where such species occur on a phylogeny. This method is useful for determining whether the state of a discrete character influences rates of evolution in another continuous character. However, many interesting evolutionary questions involve the influence of gradients, rather than discrete characters. Here, we provide a complementary approach that allows rates of evolution for a continuous trait to vary, not only between categorical states as in Brownie, but also as a function of a continuous gradient. The method tests the hypothesis that rates of evolution vary as a function of the gradient. Although models have been developed to test whether rates of molecular evolution are associated with other continuous variables such as body size (e.g. Lanfear, Welch & Bromham 2010; Lartillot & Poujol 2010), a more general modelling framework for non-molecular characters is needed.

The method we introduce uses sister-pair contrasts (sister species or other sorts of sister taxa) to estimate evolutionary rates. Sister-pair data provide a snapshot of how evolution proceeded near the tips of a phylogeny, and avoid what often are unreasonable assumptions from whole-tree approaches about how evolution may have proceeded at deeper levels in a phylogeny, where, in the absence of a detailed fossil record, trait data will be lacking. The method developed here is appropriate for data sets in which each sister pair has an estimated age of divergence (using molecular or fossil data), a Euclidean distance (or other distance metric) in a trait of interest (the trait for which evolutionary rates will be calculated), and data for a gradient along which differences in rates of evolution in the trait of interest will be tested. The method is coded into a new \textit{R} package, EVORAG: Evolutionary Rates Across Gradients. Here, we describe the modelling framework of EVORAG, test model performance, and provide advice on how to generate each of the components of a data
set. We then present a worked example in which evolutionary rates in bird song are allowed to vary between different discrete subsets (two taxonomic suborders), and within each subset, to vary as a function of a continuous environmental variable (the mid-point latitude of each sister pair).

Model description

**EvolRag** currently employs Brownian Motion (BM) and Ornstein-Uhlenbeck (OU) models of trait evolution to sister-pair data. The BM model describes a random walk in trait space through time with an instantaneous dispersion parameter, $\beta$ (i.e. the evolutionary rate, also commonly referred to as $\sigma^2$). The expectation is that a species trait value will, on average, remain the same as the starting trait value, but the variance in trait value increases as a product of time and $\beta$. This increase in variance through time is unbounded in BM (i.e. trait space has no limitations). Clearly, unbounded evolution is unrealistic for many traits, in which case an OU model may more accurately reflect evolutionary dynamics. In addition to $\beta$, the OU model adds a parameter, $\alpha$, which provides a constraint on evolution, so that trait divergence becomes more difficult as a species diverges further from a central trait value (often referred to as the optima). As $\alpha$ increases, divergence away from this central trait value becomes more difficult. $\alpha$ can thus be seen as a stabilizing selective force or as a measure of the degree of evolutionary constraint. As $\alpha$ approaches zero, the model collapses to the BM model.

Our approach uses sister-pair data by modelling trait divergence ($D$) between species within a sister pair ($i$):

$$D_i = X_{1i} - X_{2i},$$

**eqn 1**

where $X_{1i}$ and $X_{2i}$ are the mean trait values for each species in a sister pair. The set of $D_i$ is normally distributed with expectation (mean) zero and with variances (see Felsenstein 1973):

$$V_i(\text{BM} or \text{OU}) = 2\beta T_i + ME^2_{1i} + ME^2_{2i},$$

**eqn 2**

$$V_i(\text{OU}) = \frac{\beta}{\alpha}(1 - \exp(-2\alpha T_i)) + ME^2_{1i} + ME^2_{2i},$$

**eqn 3**

where $ME_{1i}$ and $ME_{2i}$ are the measurement errors (i.e. standard errors of the mean trait value estimates) for the two species in each sister pair (see eqn 1 of Felsenstein 2008). The likelihood function for BM and OU is as follows:

$$L(\beta, \alpha; D, T) = \prod \left( \frac{1}{\sqrt{2\pi V_i}} \exp \left( -\frac{D^2_i}{2V_i} \right) \right)^2,$$

**eqn 4**

This modelling framework is similar to that of Weir & Wheatcroft (2011) and Weir, Wheatcroft & Price (2012), but adds the measurement errors to the variances and corrects a minor error in their estimate of variances for BM and OU (the patterns uncovered in those papers remain valid when applying the variances in eqns 2 and 3).

The main goal of our implementation of the OU model is to allow variance to accumulate asymptotically between species pairs, regardless of the position of each species pair in phenotype space. To do this, we assume that all sister pairs have their own optimal values in trait space, which are identical to the trait values of the ancestors for each pair. Allowing each sister pair its own optimum is useful for the type of comparative analysis that evolrag is suited for, because most sister-pair data sets will pool sisters from many different clades, each of which are likely to occupy different regions of trait space. In essence, this is a model analogous to stabilizing selection, where divergence away from the ancestral trait value for each sister is penalized by $\alpha$. However, because the method implemented here models trait divergence between species rather than individual trait values for each species, the ancestor states and the optima do not appear in eqn 4 and are not estimated by the model.

The expected (i.e. mean) absolute trait divergence after a given amount of time has passed is as follows:

$$E[D|G] = \left( \frac{2|G|}{\alpha} \right)^2,$$

**eqn 5**

and follows a half normal distribution. The evolrag function expectation_time takes a set of parameters for the BM or OU models and plots the expected mean and quantiles through time. Here, we illustrate these expectations with a few examples to help compare the workings of the BM and OU models. Figure 1a plots the expectation under BM when $\beta = 1$, with data simulated under that model for comparison. At $T = 0$, the expectation of $|D|$ is 0 and then increases indefinitely through time, but not in a linear fashion. Doubling the evolutionary rate increases the expectation at any time point, but does not result in a doubling of the expectation ($\beta = 2$, Fig. 1b). This is important, because data sets with significantly different evolutionary rates may not always appear to differ greatly upon visual inspection. Under the OU model, the expectation approaches an asymptote when $\alpha > 0$ (Fig. 1c). Once the expectation has closely approached the asymptote, the expectation will not continue to increase greatly through time. The asymptote is determined by both $\alpha$ and $\beta$. When $\beta$ is held constant and $\alpha$ increases, the asymptote occurs at lower values of $|D|$ and is approached more rapidly (Fig. 1d vs. e). When $\alpha$ is held constant and $\beta$ increases, the asymptote occurs at higher values of $|D|$ (Fig. 1d vs. e).

In addition to implementing models where $\beta$ and $\alpha$ are constant across all species (BMnull and OUnull), evolrag allows $\beta$ and $\alpha$ to vary as a function of a second continuous variable, $G$, using the following three models:

1. **BMlinear**: $\beta$ changes as a linear function of $G$ ($\beta$ is replaced with $b_G G_i + c_\beta$ in eqn 1). The model has two parameters.
2. **OUlinear**: same as BMlinear but with a constant rate of $\alpha$ across the gradient. The model has three parameters.
3. **OUlinear**: both $\beta$ and $\alpha$ change as a linear function of $G$ ($\beta$ is replaced with $b_G G_i + c_\beta$ and $\alpha$ is replaced with $b_a G_i + c_\alpha$ in eqn 1). The model has four parameters.

Evolrag allows model parameters to vary as other functions of $G$ (e.g. quadratic), but these are not detailed here.

The function expectation.gradient plots the expected $|D|$ across the gradient ($G$), after a given amount of time has passed.
Fig. 1. Expectations of Euclidean distance (\( |D| \)) through time under Brownian Motion (BM) and Ornstein Uhlenbeck (OU) models for sister-pair data plotted using \textit{expectation.time}. (a) BM with \( \beta = 1 \), (b) BM with \( \beta = 2 \), (c) OU with \( \beta = 1 \) and \( \alpha = 0.1 \), (d) OU with \( \beta = 1 \) and \( \alpha = 1 \), (e) OU with \( \beta = 10 \) and \( \alpha = 1 \), (f) histograms of Euclidean distances under BM at \( T = 10 \) when \( \beta = 1 \) (light grey) and \( \beta = 4 \) (dark grey). Solid lines indicate mean, and dashed lines from bottom to top indicate the percentiles 10, 20, ..., 90, 95.

(set by the user). This function is useful for exploring the effects of the gradient on trait divergence. This is especially useful for the OU\(_{\text{linear}}\) model where change in \( \beta \) and \( \alpha \) may be positively correlated across a gradient, making it less clear whether there is an increase or decrease in expected \( |D| \) across the gradient. For example, if \( \beta \) and \( \alpha \) both increase across a gradient, the expected \( |D| \) may decline across the gradient, despite the increase in evolutionary rate. Studies using the OU\(_{\text{linear}}\) model may benefit from reporting the maximum likelihood estimate of the gradient in expected \( |D| \), rather than just \( \beta \).

**Model performance**

The simulation function in \textit{EVORAG}, \textit{sim.sisters}, uses the same approach as in packages such as geiger (Harmon et al. 2008) and diversitree (Fitzjohn 2012) to simulate trait data along the branches of a phylogeny. In our case, each simulated sister pair is represented by a phylogeny having only two tips, a single node, and with ancestor state 0. The resulting simulated trait data at the tips are then transformed into \( D \) values. Data can be simulated under the rate constant models (BM\(_{\text{null}}\) and OU\(_{\text{null}}\)) and under gradient models (BM\(_{\text{linear}}\), OU\(_{\beta\text{-linear}}\), OU\(_{\text{linear}}\)). Simulation was used to assess model performance (parameter re-estimation, model bias, type I error and statistical power) along a gradient that spanned 0–60 (as is applicable to a latitudinal gradient extending from the equator to the boreal). Six data sets were used that differed in the number of sister pairs (\( n = 30, 60, 90, 120, 150 \) and 300 sister pairs), with sisters divided evenly across five latitudes (0°, 15°, 30°, 45°, 60°) in each. Sister pairs at each latitude were divided evenly across six ages typical of sister species data sets (0.5, 1, 1.5, 2, 4, 8 million years).

**PARAMETER RE-ESTIMATION**

Simulation was used to determine whether models provided reasonable estimates of parameters. For each model, 1000 sets of \( D \) were simulated for an arbitrary trait (BM\(_{\text{null}}\): \( \beta = 0.1 \); OU-null: \( \beta = 0.1 \), \( \alpha = 0.8 \); BM\(_{\text{linear}}\): \( \beta \) at 0° = 0.1, \( \beta \) at 60° = 0.2; OU\(_{\beta\text{-linear}}\): \( \beta \) at 0° = 0.1, \( \beta \) at 60° = 0.2, \( \alpha = 0.8 \); OU\(_{\text{linear}}\): \( \beta \) at 0° = 0.1, \( \beta \) at 60° = 0.2, \( \alpha = 0.8 \), \( \alpha \) at 60° = 0.4) for each of the six data sets. The same model used to simulate trait data was then used to obtain the maximum likelihood estimate of model parameters for each simulation. All models in \textit{EVORAG} provided similar median estimates of parameter values as they were simulated under (data shown only for BM\(_{\text{linear}}\) and OU\(_{\text{linear}}\) in Fig. 2). The interquartile range of parameter re-estimates declined as the number of sis-
ters increased, indicating that larger sample sizes should result in increased accuracy of parameter estimates.

**MODEL BIAS**

To test for model bias, we simulated 1000 data sets under BM$_{null}$ with $\beta = 0.1$ for our data sets with 30 and 300 sister pairs. Simulations were then fit to the other models (BM$_{linear}$, OU$_{linear}$, OU$_{linear_beta}$), and median values of parameter estimates under these models were obtained. Bias occurs when parameters not present in BM$_{null}$ are estimated to be non-zero using the other models. No appreciable bias was found in any model parameter (Table 1). For example, the median $\alpha$ parameter of the OU$_{null}$ model was small (0.01) and slope parameters of the gradient models were very small, indicating low bias. These results contrast with those of a recent study which used similar methods to show that the $\alpha$ parameter of OU models is biased when using some whole-tree approaches (Thomas et al. 2014). While we are uncertain why some whole-tree approaches find a bias while our sister-pair approach does not, one possibility is that the need to estimate the ancestor state in whole-tree approaches, or their utilization of a single optima for all species in a tree may result in the described biases.

**RATES OF MODEL MISSPECIFICATION**

The methods in evorang are designed to test models in which rates of evolution are constant (BM$_{null}$, OU$_{null}$) vs. models in which rates vary across a gradient (BM$_{linear}$, OU$_{linear_beta}$, OU$_{linear}$). The model with the lowest AICc (Akaike information criterion adjusted for sample size) is chosen as the best fit of the candidate models to the data. We used simulation to determine rates at which a rate-variable model is incorrectly specified by AICc when the data are simulated under a rate constant model. A total of 4000 sets of D for an arbitrary trait were simulated for each of the six data sets for BM$_{null}$ ($\beta = 0.1$, and 1) and OU$_{null}$ ($\beta = 0.1$, $\alpha = 0.5$ and 2.0). Each simulated set was fit to the rate constant models (BM$_{null}$, OU$_{null}$) and to the three rate-variable models (BM$_{linear}$, OU$_{linear_beta}$, OU$_{linear}$), and AICc was used to determine whether a rate-variable or rate constant model was best supported. The proportion of simulations for which a rate-variable model was incorrectly chosen as the best fit model is shown in Fig. 3a and ranges between 0.15 and 0.2.

Models with AICc values <2 units above the best fit model ($\Delta$AICc = 2) are often considered to have substantial support along with the best supported model, while models with AICc
Table 1. Parameter bias in models simulated under BMnull with \( \beta = 0.1 \) and data set size of 30 or 300 sister pairs and fit to various test models. Median parameter values of fitted models obtained from 1000 simulations.

<table>
<thead>
<tr>
<th>( N )</th>
<th>Simulated model</th>
<th>Test model</th>
<th>Bias in ( \beta ) slope</th>
<th>Bias in ( \alpha )</th>
<th>Bias in ( \alpha ) starting</th>
<th>Bias in ( \alpha ) slope</th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td>BMnull</td>
<td>BMlinear</td>
<td>0.00002</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>300</td>
<td>BMnull</td>
<td>BMlinear</td>
<td>0.00004</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>30</td>
<td>BMnull</td>
<td>OU_null</td>
<td>NA</td>
<td>0.01326</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>300</td>
<td>BMnull</td>
<td>OU_null</td>
<td>NA</td>
<td>0.01073</td>
<td>NA</td>
<td>NA</td>
</tr>
</tbody>
</table>
| 30    | BMnull          | OU
\( f_{\text{linear}} \) | -0.00005    | 0.01076             | NA              | NA                       |
| 300   | BMnull          | OU
\( f_{\text{linear}} \) | 0.00007      | 0.00736             | NA              | NA                       |
| 30    | BMnull          | OU\( f_{\text{linear}} \) | 0.00003      | 0.00000          | 0.00000              | 0.00000              |
| 300   | BMnull          | OU\( f_{\text{linear}} \) | 0.00019      | 0.00000          | 0.00000              | 0.00000              |

BM, Brownian Motion.

Values >2 units are considered as receiving weaker support and are often not considered further (e.g., Burnham & Anderson 2001). Figure 3b shows the necessary AICc value required to maintain a model misspecification rate of 0.05. These AICc values are similar or slightly higher than the usual cut-off value of 2 units for these simulated data sets.

We also used simulation to determine rates at which a rate constant model is incorrectly specified by AICc when the data are simulated under a rate-variable model. Rates of model misspecification for rate-variable models are expected to decrease with increasing number of sister pairs in a data set and with increasing effect of the gradient on \( \beta \) and \( \alpha \). For each of the six data sets varying in the number of sister pairs, 500 sets of \( D \) for an arbitrary trait were simulated under each gradient model for the following sets of parameters. BM\( f_{\text{linear}} \), \( \beta \) at \( 0^\circ = 0.1, \beta \) at \( 60^\circ = 2, 3, 4, 6, \) and \( 10 \times \) higher than at \( 0^\circ \). OU\( f_{\text{linear}} \), the same as for BM\( f_{\text{linear}} \), but with \( \alpha = 0.5 \). OU\( f_{\text{linear}} \), four scenarios were used. Scenario 1: \( \beta \) at \( 0^\circ = 0.1, \beta \) at \( 60^\circ = 2 \times \) higher, \( \alpha \) at \( 0^\circ = 0.8, \alpha \) at \( 60^\circ = -2 \times \) lower. Scenario 2: same as scenario 1 but with \( \alpha \) at \( 60^\circ = -4 \times \). Scenario 3: same as scenario 1 but with \( \beta \) at \( 60^\circ = 4 \times \). Scenario 3: same as Scenario 3, but with \( \alpha \) at \( 60^\circ = -4 \times \). Each simulated data set was fit to each constant rate model and to each variable rate gradient model.

The proportion of simulations for which a rate constant model was incorrectly chosen as the best fit model is shown in Fig. 3c,d. As expected, model misspecification decreased with increasing values of \( n \), as well as increasing effect of the gradient on parameters. BM\( f_{\text{linear}} \) and OU\( f_{\text{linear}} \) exhibited almost identical rates of model misspecification (results are shown in Fig. 3c only for BM\( f_{\text{linear}} \), suggesting that the addition of \( \alpha \) (which is constant across the gradient in OU\( f_{\text{linear}} \)) does not affect model misspecification. For these models, a twofold increase in \( \beta \) across the gradient had low rates of model misspecification only with a data set of \( N = 300 \). Conversely, when \( \beta \) increased 10-fold, data sets with as few as 30 sisters had low rates of model misspecification. When the effect of the gradient is strong, data sets with as few as 30 sister pairs may be sufficient. When smaller gradient effects are expected, low rates of model misspecification will be retained only with larger data set sizes.

For the OU\( f_{\text{linear}} \) model, both the effect of the gradient on \( \alpha \) and \( \beta \) influences model misspecification (Fig. 3d). A twofold increase in \( \beta \) and twofold decrease in \( \alpha \) only achieved low rates of model misspecification with \( n = 90 \)–120 or higher. Both a twofold increase in \( \beta \) and fourfold decrease in \( \alpha \), or a fourfold increase in \( \beta \) and twofold decrease in \( \alpha \) had low rates of model misspecification above \( n = 60 \). A fourfold increase in \( \beta \) and fourfold decrease in \( \alpha \) had low model misspecification above \( n = 60 \), but even with \( n = 30 \), rate constant models were incorrectly favoured only in 10% of simulations. These results...
suggest that even with just fourfold changes in $\alpha$ and $\beta$, data sets as small as 30 sister pairs should provide sufficiently low rates of model misspecification.

**The software**

**INPUT DATA**

EVARAG uses sister-pair data. These may represent sister species or deeper sister clades in a phylogeny, provided sister pairs are not phylogenetically nested. The user supplies a data set of sister-pair ages ($T$), distances between species (either $D$ or $|D|$ will work in eqn 4), and values for the continuous character ($G$) over which $\beta$ and $\alpha$ will be allowed to vary. Here, we comment on strategies for measuring $T$, $D$ and $G$.

$T$: Sister-pair ages will generally be obtained from molecular data, but could also be timespans from fossil data. If age estimates are obtained from ultrametric molecular phylogenies (which do not need to be calibrated), then $T$ for each sister pair is equal to its node age. If age estimates are obtained from sequence divergence between sister pairs, then $T$ is equal to half the sequence divergence. If fossil data are used, then care must be taken that sister taxa within a sister pair are each drawn from the same time period as assumed by eqns 1–4. Likewise, branch lengths from non-ultrametric phylogenies cannot be used.

$|D|$: Euclidean distances are generally calculated on the mean or mid-point trait value for each species in a sister pair. Species mean trait values calculated from a limited sample size or with measurement error may greatly misrepresent the population mean. As a result, $|D|$ are generally biased upwards by low sample sizes. When the true distance between species is small (i.e. intraspecific variation is large relative to interspecific variation), this bias can have a strong effect on estimated Euclidean distances and could be especially important for recently diverged sister pairs (see Harmon & Losos 2005). For this reason, we strongly suggest that intraspecific sampling be maximized and standard errors of the sampling mean be obtained. For species with only a single individual measured, intraspecific variances obtained for other species in the data set can be used to estimate standard errors. These standard errors can then be incorporated into eqns 2 and 3. Standard errors assume that individuals have been sampled at random. If sampling is non-random (e.g. geographically aggregated, as is often true for museum data), then the standard error correction may not be valid. An alternative approach is to use mid-points rather than mean trait values. Provided care has been taken to sample broadly across a species geographic range (i.e. one approach is to sample populations from opposite sides of a species range), mid-point values will help avoid biases generated by aggregated sampling.

$G$: The gradient can be environmental (e.g. temperature, rainfall), geographic (e.g. latitude, longitude, elevation), ecological (e.g. net primary productivity), morphological (e.g. body mass), etc. Currently, five published studies using EVORAG have used either latitude (Weir & Wheatcroft 2011; Weir, Wheatcroft & Price 2012; Lawson & Weir 2014), the degree of climatic divergence (Lawson & Weir 2014), the degree of sexual dimorphism (Seddon et al. 2013), or relative testes size (Rowe et al. 2015) scored as a continuous character as their gradient. $G$ for each sister pair can be calculated in several ways depending on the question of interest. For the published studies using latitude, sexual dimorphism and testes size, $G$ was calculated as the absolute mid-point between the gradient scores for each individual species in a pair, under the assumption that sisters diverge in their gradient scores in a BM fashion. This approach is useful if the question of interest is to determine whether rates of evolution vary as a function of the gradient value. Because some sisters may have diverged greatly from their ancestral trait value along the gradient, it may be reasonable to include only sister pairs in which both members of a sister pair have reasonably similar gradient values. For example, for a latitudinal gradient, it makes little sense to use the mid-point latitude for a species pair in which one species occurs in the arctic and the other near the equator, because the mid-point latitude for the pair would occur in areas where neither species currently exist. For this reason, limits should be imposed on how different species within a species pair are allowed to be in their gradient values for inclusion in a data set. In the case of previous studies with latitude (and the example that follows in the next section), mid-point latitudes of species in a pair could differ by no more than 20° latitude for inclusion (Weir, Wheatcroft & Price 2012). Alternatively, if the question of interest is whether rates of evolution vary as a function of divergence in another character (i.e. an evolutionary gradient), then $G$ would be calculated as the absolute difference between the gradient values for each species (see Lawson & Weir 2014).

**THE MAXIMUM LIKELIHOOD SEARCH**

The key function in EVORAG is `model.test.sisters` which allows users to fit the above models to their data using maximum likelihood. `model.test.sisters` returns a matrix of the fitted models that includes the log-likelihood, Akaike information criterion (AIC and AICc), and parameter estimates.

**TESTING DATA SUBSETS**

Different data subsets may support different rates for the same model, or even support different models. These subsets might be defined by different taxonomic groupings, occupation of different habitats, allopatric vs. sympatric pairs, etc. To test whether a subset is significantly supported, the models in `model.test.sisters` should be fit to the entire data set without the subset and then separately to each data subset. The log-likelihood for the subsets model is simply the sum of the log-likelihood for each of the subsets, and the resulting AIC value will reflect the increased number of parameters that results from fitting the models separately to each subset. The AIC for the subsets model can then be compared to the AIC when fit to the entire data set (see Worked example below).
CONFIDENCE INTERVALS

Confidence intervals can be constructed in two ways. First, the function bootstrap.test uses bootstrap resampling of the data set to determine confidence intervals using the percentile method. Secondly, profile likelihood (Profile like.CI) can be used to calculate confidence intervals on a parameter-by-parameter basis. Both methods should generally work well. However, the likelihood surface may form a ridge between \( \alpha \) and \( \beta \) under the OU framework in some data sets, and in such cases, profile likelihood may do a better job delimiting confidence intervals when such ridges occur.

SIMULATION CAPABILITIES

EVO RAG allows data to be simulated under any of the models implemented. These can be used to test for rates of model misspecification and to determine appropriate AICc thresholds.

Worked example

Here, we illustrate the use of EVO R AG with a previously published data set on avian syllable diversity for 116 New World species pairs of passerine birds (Weir & Wheatcroft 2011). Euclidean distances (\( D \)) were derived from principal component 2 in Weir & Wheatcroft which represents an axis of syllable diversity. Species pairs with high Euclidean distance have songs with very different numbers of syllable types per song. The relative age of each sister pair (\( T \)) is estimated as half the genetic distance separating each member of the sister. Full details of this worked example are provided in Data S1.

First, we illustrate the use of EVO R AG for a discrete test of evolutionary rate differences between two major suborders of passerines: the suboscines (52 sister pairs) for which most components of song are believed to have a strong genetic component and the songbirds (64 sister pairs) for which many aspects of song (including syllable formation) have a strong culturally learned component (e.g. Baptista 1996). Song learning is generally believed to elevate rates of song evolution in songbirds, because errors in song learning and innovation can rapidly introduce new syllable types. To test these predictions, we compared rates of song learning in songbirds and suboscines using the BM_{null} models. The Akaike information criterion (AIC) indicated better support (where the model with lowest AIC value is best supported) for the model estimating separate rates for each suborder (AIC = 247.01) than for the model that estimated a single rate across all suborders (AIC = 267.74). For this best fit model, the rate measured for songbirds was four times faster than for suboscines (songbirds: \( \beta = 0.170 \); suboscines: \( \beta = 0.044 \)) leading credence to the hypothesis that learning accelerates song evolution.

However, suboscines have very low species richness at high latitudes, while songbirds have much higher species richness there. Thus, the apparent faster rates in songbirds could be an artefact of faster evolutionary rates at temperate vs. tropical latitudes, rather than between songbirds and suboscines. Many factors differ with latitude (e.g. migratory tendency, palaeoclimatic fluctuations, species richness, duration of the courting period) that could drive latitudinal differences in rate. If song learning does accelerate song evolution, then we would expect evolutionary rates to be faster in songbirds vs. suboscines at all latitudes. To test this, we included the mid-point latitude (\( G \)) of each sister pair as a continuous variable over which rates of evolution (\( \beta \)) could vary. The analysis tested for latitudinal effects in songbirds and suboscines separately and included both discrete (songbird vs. subos cine) and continuous (mid-point latitude) effects on evolutionary rates.

Adding latitude greatly improved the fit to the data (Table 2) with BM_{linear} best fit to the songbird and BM_{null} to the subos cine subsets. A total of 1000 simulations were used to determine the appropriate threshold \( \Delta \text{AICc} \) value necessary to retain a model misspecification rate of 0.05. Songbirds require a threshold \( \Delta \text{AICc} \) value of 1.7, and subs cines a value of 1.9, which are slightly lower than thresholds obtained for the simulated data sets used in the above analyses on model performance. \( \Delta \text{AICc} \) for the best fit gradient model vs. the best fit constant rate model in songbirds (12.9) greatly exceeded these thresholds indicating strong support for the gradient model, while suboscines failed to support a gradient model.

Ten thousand bootstrap replicates (using the function bootstrap.test) show the slope parameter for \( \beta \) is significantly positive in songb irds, indicating an increase in evolutionary rate with latitude (Fig. 4a). In suboscines, we also performed bootstrap replicates using BM_{linear} even though BM_{null} had best support. The maximum likelihood estimate under BM_{linear} suggests rates decline slightly with increasing latitude, although the bootstrap confidence intervals on slope indicate that this decline is not significant (Fig. 4b), thus confirming the lack of support for a gradient model. A total of 95% confidence regions for evolutionary rate were calculated across the latitudinal gradient from the bootstrap analyses (Fig. 4c).

The results show that songbirds have significantly faster evolutionary rates than suboscines only above 15° latitude and that rates for the two groups do not differ greatly at tropical latitudes. Given that song learning is believed to occur in songbirds at both tropical and temperate latitudes, these results suggest that song learning by itself cannot explain the faster rates of syllable evolution at high latitudes. Rather, some other factor that varies with latitude and interacts with song learning must drive the accelerated rates in high-latitude songbirds. Sexual selection on song is one such candidate (Irwin 2000; Weir & Wheatcroft 2011). More intense sexual selection at high latitudes (e.g. due to the short courting period and the importance of song during this period for mate choice) acting on cultural innovations in song might underlie these patterns, but requires further testing. This example illustrates how an incorrect conclusion about the positive role of song learning in accelerating evolutionary rates would have been reached if latitude had not been accounted for in this model, because without including latitude, songbirds appeared to have significantly faster evolutionary rates.
Table 2. Model fits and maximum likelihood parameter estimates for the bird song data set using the function model.test.sisters

<table>
<thead>
<tr>
<th>Model</th>
<th>logLik</th>
<th>n</th>
<th>AICc</th>
<th>ΔAICc</th>
<th>β</th>
<th>β_intercept</th>
<th>β_slope</th>
<th>α</th>
<th>α_intercept</th>
<th>α_slope</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Songbirds</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BM&lt;sub&gt;init&lt;/sub&gt;</td>
<td>-84.60</td>
<td>1</td>
<td>171.27</td>
<td>20.7</td>
<td>0.170</td>
<td>0.000</td>
<td>0.0056</td>
<td>0.744</td>
<td>0.300</td>
<td>-0.0011</td>
</tr>
<tr>
<td>BM&lt;sub&gt;linear&lt;/sub&gt;</td>
<td>-73.20</td>
<td>2</td>
<td>150.59</td>
<td>0</td>
<td>0.572</td>
<td>0.000</td>
<td>0.0010</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OU&lt;sub&gt;init&lt;/sub&gt;</td>
<td>-79.62</td>
<td>2</td>
<td>163.44</td>
<td>12.9</td>
<td>0.572</td>
<td>0.000</td>
<td>0.0010</td>
<td>0.744</td>
<td>0.300</td>
<td>-0.0011</td>
</tr>
<tr>
<td>OU&lt;sub&gt;linear&lt;/sub&gt;</td>
<td>-71.37</td>
<td>4</td>
<td>151.42</td>
<td>0.8</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.744</td>
<td>0.300</td>
<td>-0.0011</td>
</tr>
<tr>
<td><strong>Suboscines</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BM&lt;sub&gt;init&lt;/sub&gt;</td>
<td>-36.85</td>
<td>1</td>
<td>75.79</td>
<td>0</td>
<td>0.044</td>
<td>0.044</td>
<td>0.000</td>
<td>0.0715</td>
<td>0.0109</td>
<td>0.000</td>
</tr>
<tr>
<td>BM&lt;sub&gt;linear&lt;/sub&gt;</td>
<td>-36.85</td>
<td>2</td>
<td>77.95</td>
<td>2.2</td>
<td>0.054</td>
<td>0.000</td>
<td>0.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OU&lt;sub&gt;init&lt;/sub&gt;</td>
<td>-36.62</td>
<td>2</td>
<td>77.50</td>
<td>1.7</td>
<td>0.054</td>
<td>0.000</td>
<td>0.000</td>
<td>0.0715</td>
<td>0.0109</td>
<td>0.000</td>
</tr>
<tr>
<td>OU&lt;sub&gt;linear&lt;/sub&gt;</td>
<td>-34.01</td>
<td>4</td>
<td>76.87</td>
<td>1.1</td>
<td>0.129</td>
<td>-0.0017</td>
<td>0.0109</td>
<td>0.000</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\( n \), number of parameters in model; BM, Brownian Motion.
The best fit models for the songbirds and suboscines have the lowest AICc and ΔAICc (AICc – lowest AICc of candidate models) and are shown in bold.

Fig. 4. Bootstrap distributions of parameter values for best fit models (Table 2) for songbirds and suboscines. Histograms of bootstrap values for \( \beta \), slope, show that evolutionary rates increase significantly with latitude (i.e. are positive) only for songbirds. (c) The maximum likelihood estimates of \( \beta \) (thick lines) and 95% confidence region (thin lines) at each latitude for songbirds (solid lines) and suboscines (dashed lines) are plotted using the function plot.gradient.ci.

Summary

By allowing for both discrete and continuous effects on evolutionary rate, the modelling framework provided in EVORAG allows great flexibility in testing the underlying causes of trait evolution. One disadvantage of our modelling approach is that only sister-pair data are used, which results in the exclusion of some species on a phylogeny, and does not include information from interior portions of a phylogeny as in whole-tree approaches. While these models could be extended to whole-tree approaches, doing so would require ancestor state reconstruction at interior nodes, or other sorts of assumptions about how evolution proceeds at interior positions of a phylogeny. Such assumptions could compromise the validity of a whole-tree approach when testing the effects of gradients. Using sister pairs provides much greater flexibility in the modelling framework (e.g. ancestor state reconstruction is unnecessary when modelling divergences; each sister has its own optimum under OU rather than one or a few optima as in whole phylogeny approaches) and allows data to be easily pooled from unrelated taxonomic groups (e.g. for which an encompassing phylogeny is not available). For these reasons, we feel that utilization of sister-pair data is an effective means of testing the effect of continuous gradients on trait evolution.

Acknowledgements

Gavin Thomas, Trevor Price, the associate editor and one anonymous reviewer helped to improve this manuscript. This project was funded by a Natural Sciences and Engineering Research Council of Canada Discovery Grant to J.T.W. (402013-2011). Computations were performed in part on the gpc supercomputer at the SciNet HPC Consortium. SciNet is funded by: the Canada Foundation for Innovation under the auspices of Compute Canada, the Government of Ontario, Ontario Research Fund – Research Excellence and the University of Toronto.

Data accessibility

R scripts and example bird song data set: available in the R package EVORAG 2.0.

References


Received 29 May 2015; Accepted 7 July 2014

Handling Editor: Liam Revell

**Supporting Information**

Additional Supporting Information may be found in the online version of this article.

**Data S1.** EVORAG tutorial.